

Article

# Multiscale Union Regions Adaptive Sparse Representation for Hyperspectral Image Classification

Fei Tong <sup>1,2</sup> , Hengjian Tong <sup>1,\*</sup>, Junjun Jiang <sup>1</sup> and Yun Zhang <sup>2</sup>

<sup>1</sup> School of Computer Science, China University of Geosciences, Lumo Road 388, Wuhan 430074, China; tfcug93@163.com (F.T.); junjun0595@163.com (J.J.)

<sup>2</sup> Department of Geodesy and Geomatics Engineering, University of New Brunswick, 15 Dineen Drive, Fredericton, NB E3B 5A3, Canada; yunzhang@unb.ca

\* Correspondence: thj26@cug.edu.cn; Tel.: +86-027-6788-3716

Received: 13 July 2017; Accepted: 21 August 2017; Published: 23 August 2017

**Abstract:** Sparse Representation has been widely applied to classification of hyperspectral images (HSIs). Besides spectral information, the spatial context in HSIs also plays an important role in the classification. The recently published Multiscale Adaptive Sparse Representation (MASR) classifier has shown good performance in exploiting spatial information for HSI classification. But the spatial information is exploited by multiscale patches with fixed sizes of square windows. The patch can include all nearest neighbor pixels but these neighbor pixels may contain some noise pixels. Then another research proposed a Multiscale Superpixel-Based Sparse Representation (MSSR) classifier. Shape-adaptive superpixels can provide more accurate representation than patches. But it is difficult to select scales for superpixels. Therefore, inspired by the merits and demerits of multiscale patches and superpixels, we propose a novel algorithm called Multiscale Union Regions Adaptive Sparse Representation (MURASR). The union region, which is the overlap of patch and superpixel, can make full use of the advantages of both and overcome the weaknesses of each one. Experiments on several HSI datasets demonstrate that the proposed MURASR is superior to MASR and union region is better than the patch in the sparse representation.

**Keywords:** classification; hyperspectral image (HSI); multiscale union regions adaptive sparse representation (MURASR); multiscale spatial information

## 1. Introduction

Hyperspectral images have been widely applied to remote sensing image applications, such as land cover classification [1], target detection [2], anomaly detection [3], spectral unmixing [4] and others. Each pixel in HSI has hundreds of narrow contiguous bands, spanning from visible to infrared spectrum [5], which makes it possible to detect and distinguish various objects with higher accuracy [6]. However, increasing the number of spectral bands or features of an HSI pixel does not always help to increase the classification accuracy. Therefore, how to make full use of the information in HSIs is a problem in practical applications.

Many algorithms have been developed for the classification of HSIs. Among these, there are some well-known pixelwise classifiers, such as the support vector machine (SVM) [7–9], support vector conditional random classifier [10], multinomial logistic regression [11], neural network [12] and adaptive artificial immune network [13]. These pixelwise classifiers can make full use of the spectral information of HSIs, but the classification results are often noisy because the spatial information is not considered.

Therefore, some recent researches incorporated the spatial information in HSI classification to enhance the classification performance. The basic way to use spatial information is to assume

that the pixels within a local region usually represent the same material and have similar spectral characteristics [1]. Various researches [14–25] have been done based on this assumption. Besides these researches, Sparse representation (SR), which is based on the observation that spectral pixels of a particular class should lie in a low-dimensional subspace spanned by dictionary atoms (training pixels) from the same class, is also employed. In [26], a Joint Sparse Representation Classification (JSRC) method has been proposed to incorporate spectral information and spatial information. The spatial information is expressed by a fixed-size local square window centered with the test pixel. Then all pixels in the window are simultaneously joint represented by a few common atoms in the specified dictionary. The JSRC can achieve a good performance but the optimal size of the window cannot be determined easily. In [27], a stepwise Markov random field (MRF) optimization was proposed to exploit spatial information based on the result of multitask joint sparse representation. In [28], MASR was proposed to release the difficulty in choosing region size. Instead of choosing a single scale, this method extends the spatial information to several scales to take advantage of correlations among multiple region scales for HSI classification. But the multiscale regions used in MASR refer to multiscale patches which may contain noise pixels. Better than patch region, shape-adaptive superpixel can provide more accurate spatial information. In [29], the superpixel was introduced to replace the patch region. Then a shape-adaptive local smooth region was generated for each test pixel by a shape-adaptive algorithm in [30]. The latest research proposed a Multiscale Superpixel-Based Sparse Representation [31]. In this research, multiscale superpixels were generated and then each scale was represented by JSRC. Finally, a fusion result was gotten from multiscale results by majority voting. But the selection of scales for superpixels is still a problem. Although it uses multiscale to release the difficulty of selecting segmentation scale, it still needs a fundamental number of superpixels determined empirically.

In fact, patch and superpixel both have their own advantages and shortages. The patch can include all nearest neighbors but it also may contain noise pixels. Shape-adaptive superpixel can exploit more accurate spatial information but there are still some mixed superpixels when the scale is not optimal. In a mixed superpixel, there must be wrong representation because all pixels in the superpixel share the same representation. Inspired by merits and demerits of patch and superpixel, we propose to use a union region to replace the patch and superpixel. Union region refers to the overlap of patch and superpixel. Compared with patch, union region includes more similar pixels for the test pixel aiming at decreasing the effect of noise pixels. Compared with superpixel, union region provides more direct neighbors for the test pixel to enhance the representation of pixels located in the wrong superpixel. In addition, the required superpixels for generating union regions don't need empirical scale. The scales are determined by the size of the image and the corresponding patch sizes. By replacing patch in MASR with union region, we get a new algorithm called Multiscale Union Regions Adaptive Sparse Representation (MURASR). MURASR also adopts a probability majority voting method to optimize the classification result generated from the sparse representation. Experiment results show that the union region based algorithms always perform better than patch region based algorithms and the proposed MURASR outperforms other algorithms in terms of quantitative metrics and visual quality on the classification maps.

The rest parts of the paper are organized as follows. The JSRC and MASR are briefly introduced in Section 2. The details of proposed MURASR method are described in Section 3. The experimental results and discussions are presented in Section 4. Finally, Section 5 summarizes the paper and future works are suggested. The outline of the MURASR is illustrated in Figure 1.

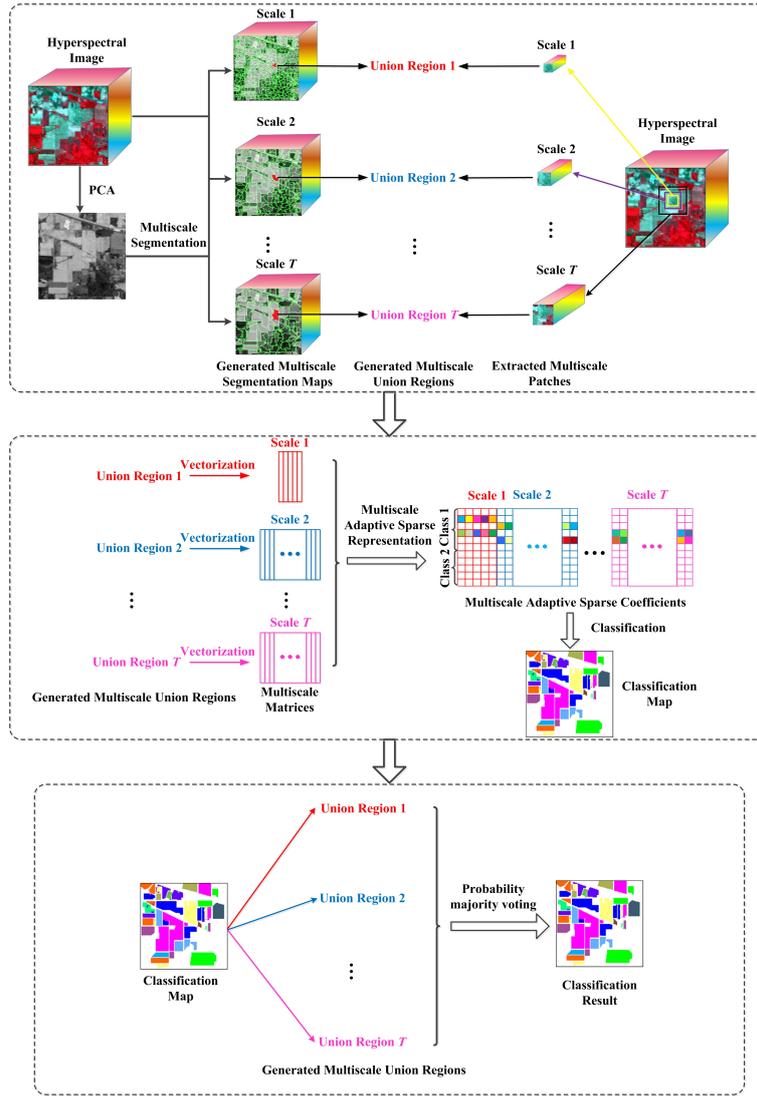


Figure 1. Outline of the proposed MURASR framework.

## 2. Background

### 2.1. JSRC

The sparse representation classification (SRC) framework was first proposed for face recognition [32]. Then Chen et al. extended the SRC to pixelwise HSI classification, which relied on the observation that spectral pixels of a particular class should lie in a low-dimensional subspace spanned by dictionary atoms (training pixels) from the same class. But spatial information is not considered by Pixelwise Sparse Representation. Therefore, based on the observation that neighboring pixels belonging to the same class usually are strongly correlated with each other, JSRC is introduced to capture such spatial correlations by assuming that neighboring pixels within a region of fixed size can be jointly represented by a few common atoms from a structural dictionary. Concretely, let  $\mathbf{y} \in \mathbb{R}^{M \times 1}$  be a pixel with  $M$  denoting the number of spectral bands and  $\mathbf{D} = [\mathbf{D}_1, \dots, \mathbf{D}_c, \dots, \mathbf{D}_C] \in \mathbb{R}^{M \times N}$  be a structure dictionary, where  $\mathbf{D}_c \in \mathbb{R}^{M \times N_c}$ ,  $c = 1, \dots, C$  is the  $c$ th class subdictionary whose columns (atoms) are extracted from the training pixels;  $C$  is the number of classes;  $N_c$  is the number of atoms in subdictionary  $\mathbf{D}_c$ ; and  $N = \sum_{c=1}^C N_c$  is the total number of atoms in  $\mathbf{D}$ . Specifically, the size of a region

surrounding the test pixel  $\mathbf{y}_1$  is denoted by  $W \times W$ , and pixels within such a region can be denoted by a matrix  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{W \times W}]$ . The matrix can be compactly represented as:

$$\begin{aligned} \mathbf{Y} &= [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{W \times W}] = [\mathbf{D}\mathbf{A}_1, \mathbf{D}\mathbf{A}_2, \dots, \mathbf{D}\mathbf{A}_{W \times W}] \\ &= \mathbf{D}[\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{W \times W}] = \mathbf{D}\mathbf{A} \end{aligned} \quad (1)$$

where  $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{W \times W}]$  is the sparse coefficients matrix corresponding to  $\mathbf{Y}$ . Since the indexes of the selected atoms in  $\mathbf{D}$  are determined by the positions of nonzero coefficients in  $[\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{W \times W}]$ , the neighboring pixels  $[\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{W \times W}]$  can be represented by a small set of common atoms by enforcing a few nonzero rows on the sparse coefficients matrix  $\mathbf{A}$ . Then, matrix  $\mathbf{A}$  can be obtained by solving the following optimization problem:

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \|\mathbf{Y} - \mathbf{D}\mathbf{A}\|_F \quad \text{subject to } \|\mathbf{A}\|_{\text{row},0} \leq K \quad (2)$$

where  $\mathbf{A}_{\text{row},0}$  denotes the joint sparse norm, which is used to select a number of the most representative nonzero rows in  $\mathbf{A}$ , and  $\|\cdot\|_F$  is the Frobenius norm. A variant of the OMP algorithm called the simultaneous OMP (SOMP) [33,34], can be used to efficiently obtain an approximate solution. After  $\hat{\mathbf{A}}$  is recovered, the label of test pixel  $\mathbf{y}_1$  can be decided by the minimal total error:

$$\hat{c} = \arg \min_c \|\mathbf{Y} - \mathbf{D}_c \hat{\mathbf{A}}_c\|_F, \quad c = 1, \dots, C \quad (3)$$

where  $\hat{\mathbf{A}}_c$  denotes the rows in  $\hat{\mathbf{A}}$  associated with the  $c$ th class.

## 2.2. MASR

Compared with pixelwise SRC model, the JSRC can achieve more accurate classification results because of incorporating spatial information of local regions. However, the region size (or the *region scale*) has great influence on the classification performance. It is of great importance to determine an optimal region scale for the JSRC.

Then Fang et al. proposed the MASR to release the difficulty of choosing region scale. The MASR effectively exploits spatial information at multiple scales via an adaptive sparse strategy. Not only does the adaptive sparse strategy restrict pixels from different scales to be represented by training atoms from a particular class but also allow the selected atoms for these pixels to be varied, thus providing an improved representation. Given one test pixel  $\mathbf{y}_1$  in HSI, its  $T$  neighboring regions are selected via different predefined scales. Neighboring regions are defined by multiscale patches centered with test pixel. Then a multiscale matrix  $\mathbf{Y}^{\text{mp}} = [\mathbf{Y}_1, \dots, \mathbf{Y}_t, \dots, \mathbf{Y}_T]$  can be constructed by pixels within the selected regions, where the  $\mathbf{Y}_t$  includes pixels from the  $t$ th scale region. Since spatial structures and characteristics for different scales of regions are distinct, the generated multiscale matrix  $\mathbf{Y}^{\text{mp}}$  for the test pixel  $\mathbf{y}_1$  should provide complementary yet correlated information, which can be utilized to classify  $\mathbf{y}_1$  more accurately.

In MASR, an adaptive sparse strategy is adopted to utilize the correlated information among multiscales and achieve a flexible selection process for atoms. An important part of the adaptive strategy is the adoption of a collection of adaptive sets. Each adaptive set is denoted as the indexes of a set of nonzero scalar coefficients, which belong to the same class in the multiscale sparse matrix  $\mathbf{A}^{\text{mp}}$ . By combining the adaptive set with the  $\ell_{\text{row},0}$  norm, a new adaptive norm  $\ell_{\text{adaptive},0}$  is created on  $\mathbf{A}^{\text{mp}}$ , which can be used to select a small number of adaptive sets from  $\mathbf{A}^{\text{mp}}$ . Then,  $\mathbf{A}^{\text{mp}}$  matrix can be recovered by applying the adaptive norm as follows:

$$\begin{aligned} \hat{\mathbf{A}}^{\text{mp}} &= \arg \min_{\mathbf{A}^{\text{mp}}} \|\mathbf{Y}^{\text{mp}} - \mathbf{D}\mathbf{A}^{\text{mp}}\|_F \\ &\text{subject to } \|\mathbf{A}^{\text{mp}}\|_{\text{adaptive},0} \leq K \end{aligned} \quad (4)$$

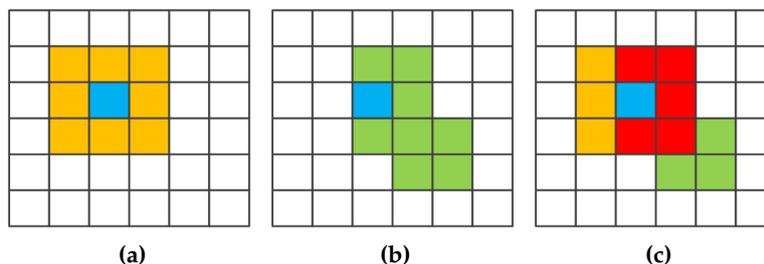
After recovering the multiscale sparse representation matrix  $\hat{\mathbf{A}}^{\text{mp}}$ , a single decision can be made on the test pixel  $\mathbf{y}_1$  based on the lowest total representation error:

$$\hat{c} = \arg \min_c \|\mathbf{Y}^{\text{mp}} - \mathbf{D}_c \hat{\mathbf{A}}_c^{\text{mp}}\|_F, c = 1, \dots, C \quad (5)$$

where  $\hat{\mathbf{A}}_c^{\text{mp}}$  represents rows in  $\hat{\mathbf{A}}^{\text{mp}}$  corresponding to the  $c$ th class.

### 3. Multiscale Union Regions Adaptive Sparse Representation

The aforementioned MASR shows good performance for HSI classification. But the MASR utilizes multiscale patches to exploit spatial information. In a patch, maybe most of the pixels are different from the test pixel, such as a pixel on the edge of a building. The classification may be misled by those noise pixels from other classes which are similar to the atoms in the dictionary, thus providing an incorrect classification for test pixel. In computer vision, superpixels have been studied to provide an efficient representation, which can facilitate visual recognition [35–37]. Each superpixel is a perceptually meaningful region, whose shape and size can be adaptively changed according to different spatial structures. But how to find an optimal scale for superpixels is still a challenge. Without optimal scale, some mixed superpixels will be generated. Based on the fact that patch and superpixel may include pixels from different classes, a multiscale union regions adaptive sparse representation model is proposed to decrease the influence of noise pixels for the test pixel. The union region is the overlap of the patch and corresponding superpixel with the same scale (see Figure 2). For a test pixel, if the patch includes some noise pixels, the superpixel can provide more similar pixels to reduce the impact of noise pixels. In the same way, if the test pixel is located in the wrong superpixel which has seldom pixels similar to test pixel, the patch can provide more similar pixels to enhance the right representation.



**Figure 2.** Three kinds of spatial regions: (a) fixed-size patch; (b) adaptive size superpixel; and (c) union of patch and superpixel. The blue pixel represents test pixel, orange pixels are neighbors defined by patch, green pixels are neighbors defined by superpixel and red pixels are overlap of neighbors defined by patch and superpixel.

#### 3.1. Generation of Multiscale Union Regions

Before generating multiscale union regions, we should get multiscale superpixels. There are various researches focusing on the segmentation [36–39]. In this paper, an oversegmentation algorithm called ERS [37] is applied to generate 2-D superpixel maps on the base images because of its high efficiency. Unlike the single-band gray or three-band color image, the HSI usually has hundreds of spectral bands. To improve the computational efficiency, PCA [40] is first used to reduce the spectral bands of the HSI. Since the important information of the HSI exists in the principle components (e.g., first three principle components), they are used as the base images. In this paper, only the first principle component is chosen as the base image. Instead of choosing scales for superpixels empirically, we calculate scales of superpixels based on corresponding patch sizes. Assuming that  $\mathbf{PS}_i$  refers to the

patch size of  $t$ th scale and  $N_{total}$  is the total number of pixels in the image (note that origin image will be extended for edge pixels), the superpixels number  $n_t$  for  $t$ th segmentation is calculated as:

$$n_t = N_{total} / PS_t \tag{6}$$

In this way, the average size of superpixels is equal to patch size. Then most superpixels will have similar sizes with patches. It guarantees that superpixel and patch can have similar influence on union region. What's more, with the increasing of patch size, the superpixels number decreases fast. Thus, only limited number of segmentations can be generated. According to the performance of limited number of segmentations, it will be easier for users to determine the scales number. After segmentations,  $T$  superpixels are generated for each test pixel  $y_1$  and these superpixels construct the corresponding multiscale matrix  $Y^{ms} = [Y_1, \dots, Y_t, \dots, Y_T]$ , where the  $Y_t$  includes pixels from the  $t$ th superpixel. Then for a specific  $t$ th scale, the union region  $Y_t^{mu}$  is defined as following:

$$Y_t^{mu} = Y_t^{ms} \cup Y_t^{mp} \tag{7}$$

### 3.2. Multiscale Union Regions Adaptive Sparse Representation

For a test pixel  $y_1$ , the corresponding multiscale matrix is  $Y^{mu} = [Y_1, \dots, Y_t, \dots, Y_T]$ , where  $Y_t$  is the union of  $Y_t^{mp}$  and  $Y_t^{ms}$ . Then the sparse coefficients matrix  $A^{mu}$  can be recovered by solving following problem:

$$\begin{aligned} \hat{A}^{mu} &= \arg \min_{A^{mu}} \|Y^{mu} - DA^{mu}\|_F, \\ &\text{subject to } \|A^{mu}\|_{\text{adaptive},0} \leq K \end{aligned} \tag{8}$$

To solve this problem, the method used in MASR is applied. At each iteration, the current residual correlation matrix is calculated firstly. Then a new adaptive set based on the current residual correlation matrix will be selected. Once the selecting of the new adaptive set is finished, the new adaptive set will be merged with previously selected adaptive sets. Then the sparse coefficients matrix is estimated based on the merged adaptive sets. Finally, the residue is updated. The iterations will stop if the termination criterion is satisfied. After the multiscale sparse representation matrix  $\hat{A}^{mu}$  is recovered, the final label of the test pixel  $y_1$  can be determined by minimal total representation error:

$$\hat{c} = \arg \min_c \|Y^{mu} - D_c \hat{A}_c^{mu}\|_F, c = 1, \dots, C \tag{9}$$

### 3.3. Probability Majority Voting

Because multiscale union regions adaptive sparse representation is a pixel-based classifier, there will be some pepper salt noise pixels in ground truth objects. Therefore, a majority voting process will be helpful to optimize the classification result. As mentioned above, for each test pixel in each scale, a union region will be generated. Then for the union region, the probabilities belonging to all classes are calculated. If a union region at  $i$ th scale contains  $N_i^{total}$  labeled pixels and  $N_i^j$  pixels classified to  $j$ th class, the probability belonging to  $j$ th class  $P_i^j$  is calculated as:

$$P_i^j = N_i^j / N_i^{total} \tag{10}$$

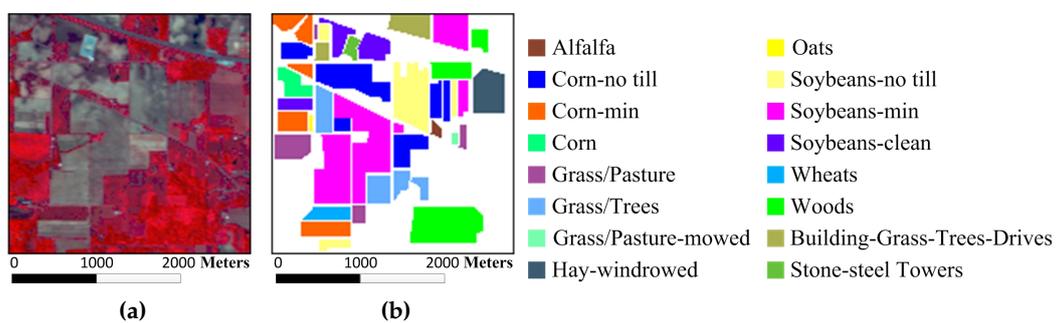
Assuming that there are  $k$  classes,  $T$  scales of segmentation maps, the class label of the test pixel  $\hat{j}$  can be obtained by:

$$\hat{j} = \arg \max_j \left( \sum_{i=1}^T P_i^j \right), j = 1, \dots, k \tag{11}$$

## 4. Experimental Results and Discussion

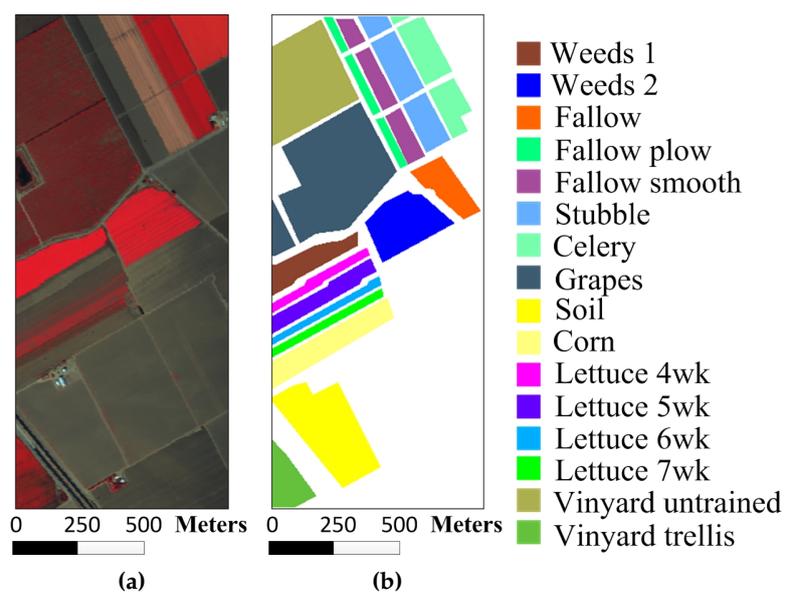
### 4.1. Data Sets

To verify the effectiveness of the proposed MURASR method and superiority of the union region, experiments are conducted on the following three hyperspectral data sets: the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian Pines data, the AVIRIS Salinas data, and the Reflective Optics System Imaging Spectrometer (ROSIS-03) University of Pavia data. The AVIRIS Indian Pines image has 220 data channels with the size of  $145 \times 145$  across the spectral range from 0.2 to  $2.4 \mu\text{m}$ . It was captured over the agricultural Indian Pine test site in northwestern Indiana with a spatial resolution of 20 m per pixel. Before classification, 20 water absorption bands (No. 104–108, 150–163 and 220) were discarded [41]. Figure 3a,b show the color composite of the Indian Pines image and the corresponding reference data with 16 reference classes from different types of crops.



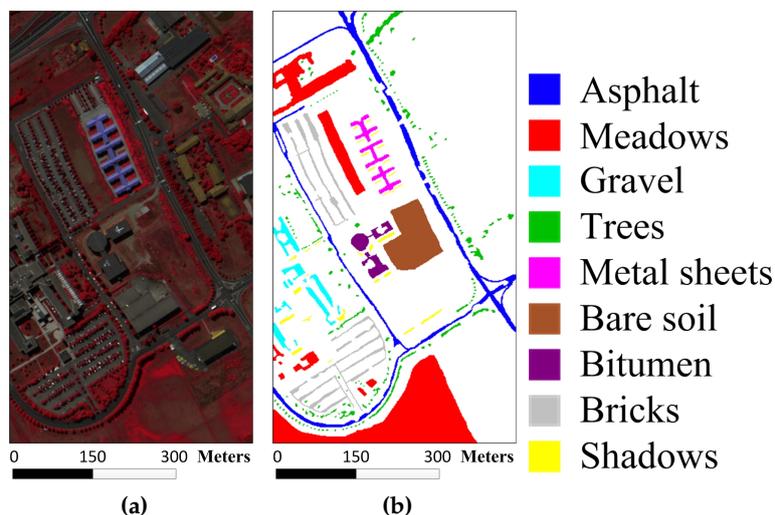
**Figure 3.** Indian Pines image: (a) three-band color composite image; (b) reference image.

The Salinas image was also acquired by the AVIRIS sensor over Salinas Valley, California. The image is of size  $512 \times 217 \times 224$  with a spatial resolution of 3.7 m per pixel. Similar to the Indian Pines image, 20 water absorption spectral bands (No. 108–112, 154–167 and 224) were removed and 16 different reference classes are considered for this image. Figure 4a,b show the color composite of the Salinas image and the corresponding reference data.



**Figure 4.** Salinas image: (a) three-band color composite image; (b) reference image.

The University of Pavia image, which captures an urban area surrounding the University of Pavia, Italy, was recorded by the ROSIS-03 sensor. The image is of size  $610 \times 340 \times 115$  with a spatial resolution of 1.3 m per pixel and a spectral coverage ranging from 0.43 to  $0.86 \mu\text{m}$ . The 12 very noisy channels were discarded before the experiments, and nine information classes are considered for this image. Figure 5a,b show the color composite of the University of Pavia image and the corresponding reference data.



**Figure 5.** University of Pavia image: (a) three-band color composite image; (b) reference image.

#### 4.2. Comparison of Experiment Results

In the experiments, all related algorithms are based on sparse representation. Except for published algorithms SRC, JSRC and MASR, JUSRC (Joint Union Sparse Representation Classification), MJSRC (Multiscale Joint Sparse Representation Classification), MJUSRC (Multiscale Joint Union Sparse Representation Classification), MURASR\* and MURASR were conducted in the experiments. To verify the priority of union region further, the patch used in JSRC was replaced by JUSRC with the union region. For demonstrating the superiority of multiscale adaptive strategy, we extended the JSRC and JUSRC with a simple multiscale scheme that applied the majority voting to the results of all scales for the final decision-making. The extended algorithms are called MJSRC and MJUSRC. What's more, the MURASR\* is the MURASR without probability majority voting process. The comparison between MURASR\* and MURASR can show the difference of whether the probability majority voting method was used or not. The parameters for the SRC, JSRC, and JUSRC algorithms were tuned to reach the best results in these experiments. For all multiscale algorithms, seven different scales were simultaneously adopted, and the selected region scales were as follows:  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ ,  $11 \times 11$ ,  $13 \times 13$ , and  $15 \times 15$ . Then superpixels numbers for segmentation were calculated with Equation (6) and listed in Table 1. Other parameters in MJSRC, MJUSRC, MASR, MURASR\*, and MURASR were the same as [28]. To evaluate the performance of classifiers, three objective metrics (overall accuracy (OA), average accuracy (AA) and kappa coefficient) are adopted. In addition, the McNemar's test is applied to analyse the experiment results. The McNemar's test is based on the standardized normal test statistic, as described in [42]:

$$\mathbf{Z} = \frac{h_{12} - h_{21}}{\sqrt{h_{12} + h_{21}}} \quad (12)$$

where  $h_{12}$  represents the samples correctly classified by method 1 but incorrectly classified by method 2. If  $|\mathbf{Z}| > 1.96$ , the accuracy between two methods can be considered statistically significant. The sign of the  $\mathbf{Z}$  indicates which method is better. If  $\mathbf{Z} > 0$ , the method 1 is more accurate than method 2.

The Indian Pines data set was classified firstly. 10% of the labeled pixels were randomly sampled for training from each class, while the rest 90% were used to test the classifiers (see Table 2). The classification maps generated by different classifiers on the Indian Pines image are shown in Figure 6. The details of the classification results averaged by ten runs with randomly sampled training samples are tabulated in Table 3. The results of the McNemar's tests between classifiers are listed in Table 4. It is easy to find that JUSRC, MJUSRC and MURASR\* perform better than JSRC, MJSRC and MASR, which demonstrates the priority of union region over patch region. In addition, the multiscale majority voting based MJSRC and MJUSRC perform worse than the multiscale adaptive strategy based MASR and MURASR\* for this image. Compared with MJSRC and MJUSRC, accuracy improvements of MASR and MURASR\* are more than 3%. MURASR gets a better result than MURASR\* in accuracy and classification map. As can be observed from the classification maps of MURASR\* and MURASR, many misclassifications in MURASR\* can be eliminated efficiently by probability majority voting method. What's more, MURASR performs best among all algorithms in terms of OA and AA, and the results of the McNemar's test are statistically significant and coherent with the obtained overall accuracies.

**Table 1.** Number of Superpixels in Each Scale.

	1	2	3	4	5	6	7
Indian Pines	2809	1011	515	312	208	149	112
Salinas	13,500	4860	2479	1500	1004	718	540
University of Pavia	24,544	8835	4508	2727	1825	1307	981

**Table 2.** Sixteen reference classes in the Indian Pines image.

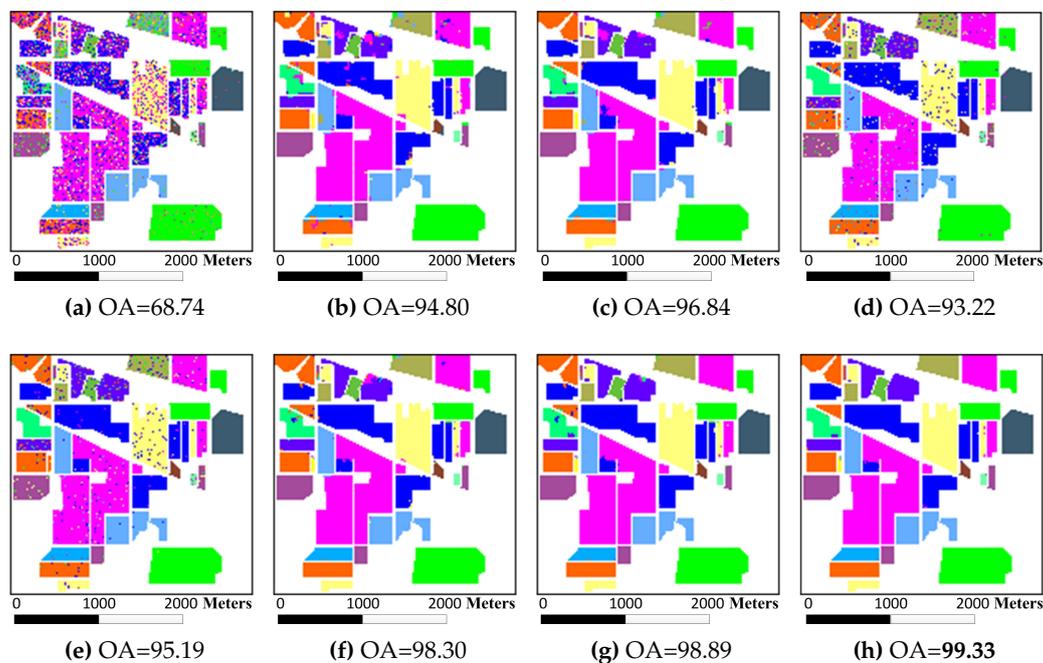
Class	Name	Train	Test
1	Alfalfa	5	41
2	Corn-no till	143	1285
3	Corn-min	83	747
4	Corn	24	213
5	Grass/Pasture	48	435
6	Grass/Trees	73	657
7	Grass/Pasture-mowed	3	25
8	Hay-windrowed	48	430
9	Oats	2	18
10	Soybeans-no till	97	875
11	Soybeans-min	246	2209
12	Soybeans-clean	59	534
13	Wheat	21	184
14	Woods	127	1138
15	Building-Grass-Trees-Drives	39	347
16	Stone-steel Towers	9	84
Total		1027	9222

The second experiment was performed on the Salinas data set. To compare the classification with MASR, only 1% of the labeled pixels for each class were randomly selected for training. Then the remaining 99% labeled data were classified with the classifiers to demonstrate the superiority of the proposed MURASR (see Table 5). The classification maps for various classifiers are illustrated in Figure 7 and the average quantitative results of ten runs are tabulated in Table 6. Moreover, the results of the McNemar's tests are shown in Table 7. As can be observed, union region based algorithms JUSRC, MJUSRC and MURASR\* still get more accurate results than patch region based JSRC, MJSRC and MASR in terms of OA, AA and Kappa coefficients. The classification maps of MJSRC and MJUSRC have more pepper salt noise pixels than MASR and MURASR\*. Comparing classification maps of

MURASR\* and MURASR, we can find that most misclassifications generated by MURASR\* can be corrected by probability majority voting method. In addition, the average accuracy of MURASR is 99.70% which is very high for classification. Moreover, it should be noted that the McNemar's tests between classifiers are also statistically significant and coherent with the obtained overall accuracies.

The final experiment was conducted on the University of Pavia image. The shapes of surface objects in this image are more complex than previous two images. For each reference class, 200 train samples were randomly selected from the labeled data and the remaining pixels were used for testing the performance of various classifiers (see Table 8). The classification maps are demonstrated in Figure 8 and the detail data averaged by ten runs in term of OA, AA, and Kappa coefficients is listed in Table 9. The McNemar's tests between classifiers also were conducted on this image and the results are tabulated in Table 10. Same as previously mentioned two images, union region based classifiers also performed better than patch region based classifiers. Multiscale adaptive strategy still works better than multiscale majority voting strategy in this image. The accuracy improvement gained by probability majority voting is less than previous two images because the University of Pavia image has less large homogenous regions. And from Table 9, we can find that MASR only has more accurate result than MURASR with one class and MURASR performs best among all classifiers with 7 classes, which proves the priority of MURASR further. The results of the McNemar's tests also provide enough support for the analysis.

Compared with many presented algorithms, MASR is a time-consuming algorithm. In this paper, the proposed MURASR is designed based on the multiscale adaptive representation in MASR. Also, the generation of union regions will consume some time. Moreover, the union region has more pixels than patch region. Therefore, the MURASR is also a time-consuming algorithm and the time cost of MURASR is about twice as much as MASR. But the proposed MURASR was coded in MATLAB (R2016a, Mathworks, Portola Valley, CA, USA) and was not optimized for speed. The MURASR can be significantly sped up by changing the compiling code from MATLAB to C++ and adopting a general-purpose graphics processing unit (GPU).



**Figure 6.** Classification maps for the Indian Pines image by different algorithms: (a) SRC-Pixel-Wise; (b) JSRC; (c) JUSRC; (d) MJSRC; (e) MJUSRC; (f) MASR; (g) MURASR\*; and (h) MURASR.

**Table 3.** Classification accuracy (averaged on ten runs with randomly sampled training samples) of the Indian Pines image. The best results are highlighted in bold typeface.

Class	SRC-Pixel-Wise	JSRC	JUSRC	MJSRC	MJUSRC	MASR	MURASR*	MURASR
1	35.12	87.56	96.83	95.37	96.34	93.66	96.83	<b>98.54</b>
2	54.63	94.87	96.48	94.39	95.25	97.77	<b>97.93</b>	97.84
3	51.99	93.44	97.00	91.67	95.69	98.17	98.77	<b>99.54</b>
4	36.53	89.62	95.31	91.50	92.77	94.89	95.77	<b>98.78</b>
5	82.44	94.28	95.38	92.11	93.17	95.59	96.23	<b>96.51</b>
6	93.32	97.43	98.95	96.19	98.42	99.83	<b>100</b>	<b>100</b>
7	66.80	96.80	94.40	66.40	66.80	<b>98.80</b>	98.40	96.00
8	95.93	99.44	99.79	98.60	98.70	99.95	99.98	<b>100</b>
9	17.78	60.56	<b>91.11</b>	12.22	19.44	64.44	79.44	71.67
10	65.99	95.67	97.58	89.14	91.55	97.68	<b>98.23</b>	97.67
11	71.52	96.68	97.78	95.91	95.76	99.01	99.11	<b>99.85</b>
12	41.82	89.76	95.30	87.83	92.47	96.55	98.15	<b>99.25</b>
13	92.28	94.95	98.37	90.43	97.83	98.75	99.29	<b>99.89</b>
14	88.93	98.99	99.33	99.24	99.92	99.95	99.96	<b>100</b>
15	35.45	89.05	93.83	92.54	91.84	97.52	98.70	<b>99.48</b>
16	89.40	88.33	92.02	81.90	89.40	96.07	96.90	<b>98.69</b>
OA	68.83	95.35	97.36	93.91	95.35	98.29	98.69	<b>99.06</b>
AA	64.40	94.69	96.98	93.06	94.71	95.55	97.11	<b>98.93</b>
Kappa	0.64	0.92	0.96	0.86	0.88	0.98	<b>0.99</b>	0.97

**Table 4.** The McNemar's tests between classifiers (averaged on ten runs with randomly sampled training samples) of the Indian Pines image.

Method	JSRC	JUSRC	MJSRC	MJUSRC	MASR	MURASR*	MURASR
<b>JSRC</b>	–	–10.56	5.24	–0.25	–14.19	–15.28	–16.33
<b>JUSRC</b>	10.56	–	13.34	8.37	–5.15	–8.27	–10.09
<b>MJSRC</b>	–5.24	–13.34	–	–7.63	–17.24	–18.77	–20.26
<b>MJUSRC</b>	0.25	–8.37	7.63	–	–12.29	–14.84	–17.17
<b>MASR</b>	14.19	5.15	17.24	12.29	–	–3.51	–5.65
<b>MURASR*</b>	15.28	8.27	18.77	14.84	3.51	–	–3.40
<b>MURASR</b>	16.33	10.09	20.26	17.17	5.65	3.40	–

**Table 5.** Sixteen reference classes in the Salinas image.

Class	Name	Train	Test
1	Weeds_1	20	1989
2	Weeds_2	37	3689
3	Fallow	20	1956
4	Fallow plow	14	1380
5	Fallow smooth	27	2651
6	Stubble	40	3919
7	Celery	36	3543
8	Grapes	113	11,158
9	Soil	62	6141
10	Corn	33	3245
11	Lettuce 4 wk	11	1057
12	Lettuce 5 wk	19	1908
13	Lettuce 6 wk	9	907
14	Lettuce 7 wk	11	1059
15	Vinyard untrained	73	7195
16	Vinyard trellis	18	1789
	Total	543	53,586

**Table 6.** Classification accuracy (averaged on ten runs with randomly sampled training samples) of the Salinas image. The best results are highlighted in bold typeface.

Class	SRC-Pixel-Wise	JSRC	JUSRC	MJSRC	MJUSRC	MASR	MURASR*	MURASR
1	98.23	<b>100</b>	<b>100</b>	99.99	<b>100</b>	99.98	<b>100</b>	<b>100</b>
2	98.04	99.95	<b>100</b>	99.95	99.94	99.78	99.79	<b>100</b>
3	94.16	99.33	99.71	99.07	99.68	99.38	99.86	<b>100</b>
4	98.77	70.59	87.10	85.46	94.91	97.31	98.83	<b>99.88</b>
5	91.84	85.98	92.08	93.77	97.71	99.07	99.51	<b>99.52</b>
6	99.41	95.68	96.81	99.26	99.57	<b>100</b>	<b>100</b>	<b>100</b>
7	99.16	97.65	98.49	99.57	99.72	99.95	99.92	<b>100</b>
8	70.99	95.19	98.07	94.29	96.52	96.41	98.39	<b>99.61</b>
9	97.23	99.98	99.99	99.97	<b>100</b>	99.91	99.95	<b>100</b>
10	85.45	93.78	95.69	96.76	97.63	98.06	98.44	<b>99.63</b>
11	93.56	88.91	93.81	98.34	99.13	99.91	99.92	<b>100</b>
12	99.75	88.68	94.33	96.16	98.95	99.85	99.96	<b>100</b>
13	97.14	81.52	89.35	95.64	97.65	99.26	99.46	<b>99.98</b>
14	92.64	85.15	87.18	96.95	97.56	98.59	98.53	<b>99.93</b>
15	59.14	91.69	96.03	87.90	92.44	93.12	96.74	<b>98.79</b>
16	93.93	99.73	99.55	99.65	99.64	99.16	99.14	<b>99.78</b>
OA	85.79	94.32	96.96	95.87	97.65	97.97	98.98	<b>99.70</b>
AA	84.19	93.67	96.62	95.40	97.38	98.73	99.28	<b>99.66</b>
Kappa	0.92	0.92	0.96	0.96	0.98	0.98	0.99	<b>1</b>

**Table 7.** The McNemar's tests between classifiers (averaged on ten runs with randomly sampled training samples) of the Salinas image.

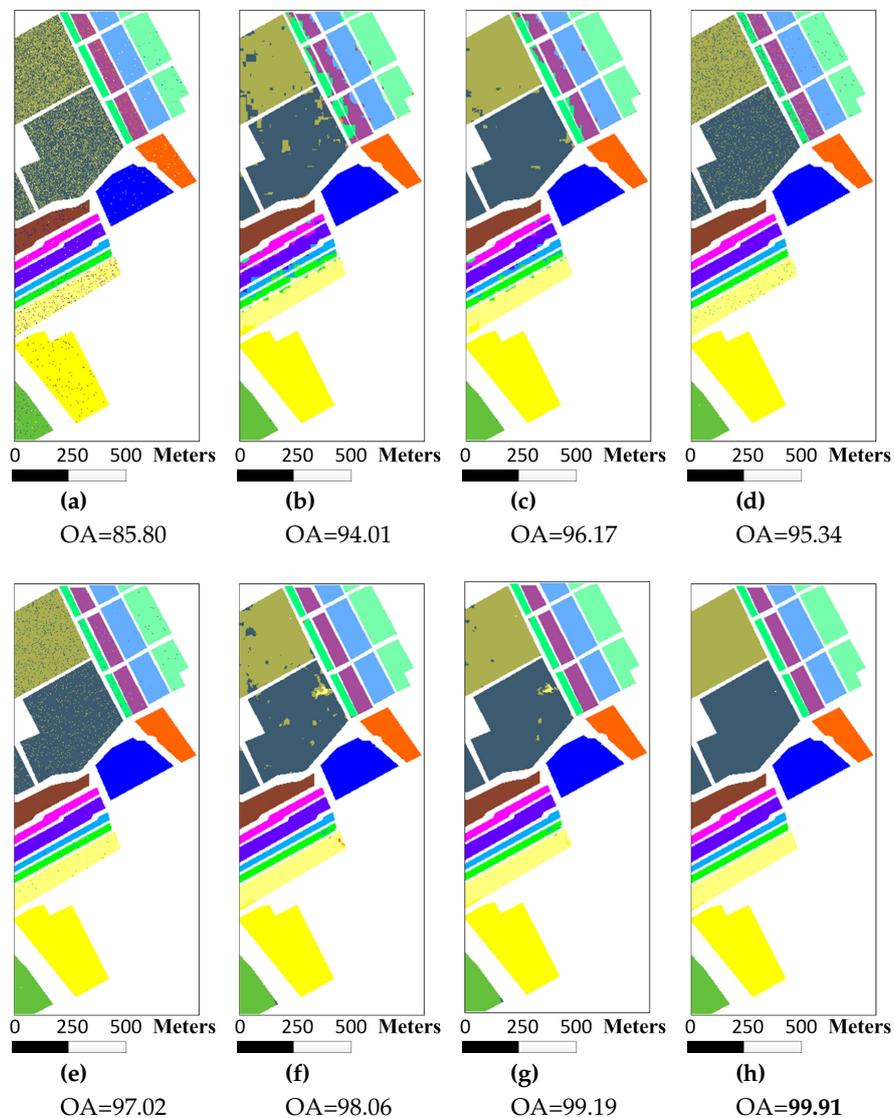
Method	JSRC	JUSRC	MJSRC	MJUSRC	MASR	MURASR*	MURASR
<b>JSRC</b>	–	–32.24	–17.31	–34.78	–37.06	–46.17	–53.08
<b>JUSRC</b>	32.24	–	11.43	–9.24	–11.66	–26.65	–36.90
<b>MJSRC</b>	17.31	–11.43	–	–26.12	–28.24	–37.55	–44.79
<b>MJUSRC</b>	34.78	9.24	26.12	–	–4.82	–22.67	–32.56
<b>MASR</b>	37.06	11.66	28.24	4.82	–	–20.12	–29.66
<b>MURASR*</b>	46.17	26.65	37.55	22.67	20.12	–	–18.89
<b>MURASR</b>	53.08	36.90	44.79	32.56	29.66	18.89	–

**Table 8.** Nine reference classes in the University of Pavia image.

Class	Name	Train	Test
1	Asphalt	200	6431
2	Meadows	200	18,449
3	Gravel	200	1899
4	Trees	200	2864
5	Metal sheets	200	1145
6	Bare soil	200	4829
7	Bitumen	200	1130
8	Bricks	200	3482
9	Shadows	200	747
	Total	1800	40,976

**Table 9.** Classification accuracy (averaged on ten runs with randomly sampled training samples) of the University of Pavia image. The best results are highlighted in bold typeface.

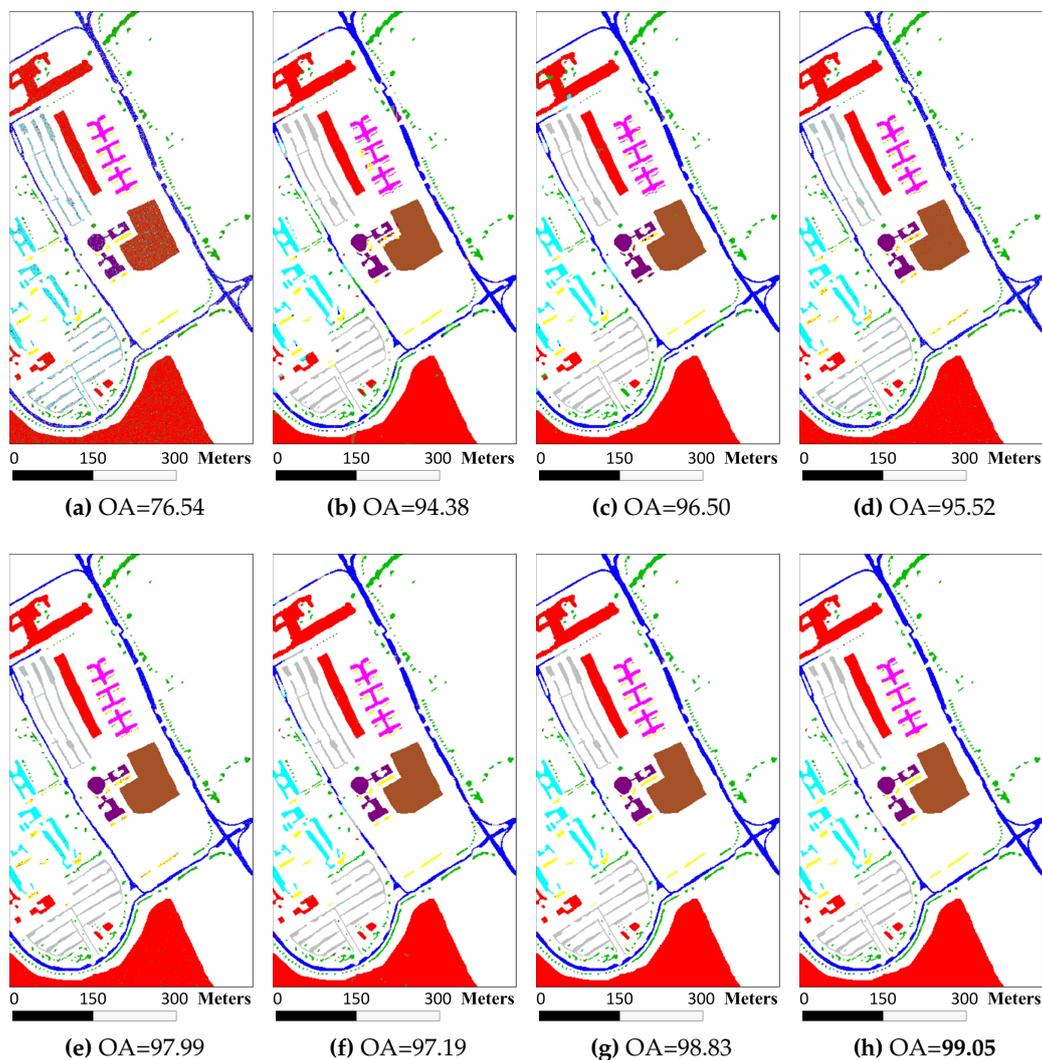
Class	SRC-Pixel-Wise	JSRC	JUSRC	MJSRC	MJUSRC	MASR	MURASR*	MURASR
1	62.24	86.22	93.13	86.14	94.79	89.97	96.87	<b>98.38</b>
2	80.22	96.62	97.15	97.71	98.63	98.78	99.44	<b>99.70</b>
3	69.07	98.64	99.36	99.27	99.76	99.78	99.87	<b>99.89</b>
4	91.48	91.32	90.94	95.89	94.99	<b>97.47</b>	97.13	95.66
5	99.52	97.69	99.11	99.55	99.78	<b>100</b>	<b>100</b>	<b>100</b>
6	68.84	99.54	99.71	99.53	99.93	99.87	99.93	<b>100</b>
7	86.90	97.53	99.59	99.79	99.92	<b>100</b>	<b>100</b>	<b>100</b>
8	72.67	95.89	97.92	96.70	98.68	98.76	99.72	<b>99.94</b>
9	<b>98.17</b>	66.96	75.17	84.08	86.43	92.16	94.90	95.30
OA	76.74	94.51	96.27	95.83	97.83	97.42	98.92	<b>99.21</b>
AA	69.61	92.67	95.02	94.42	97.09	96.54	98.55	<b>98.94</b>
Kappa	0.81	0.92	0.95	0.95	0.97	0.97	<b>0.99</b>	<b>0.99</b>



**Figure 7.** Classification maps for the Salinas image by different algorithms: (a) SRC-Pixel-Wise; (b) JSRC; (c) MUSRC; (d) MJUSRC; (e) MASR; (f) MURASR\*; (g) MURASR; and (h) MURASR.

**Table 10.** The McNemar’s tests between classifiers (averaged on ten runs with randomly sampled training samples) of the University of Pavia image.

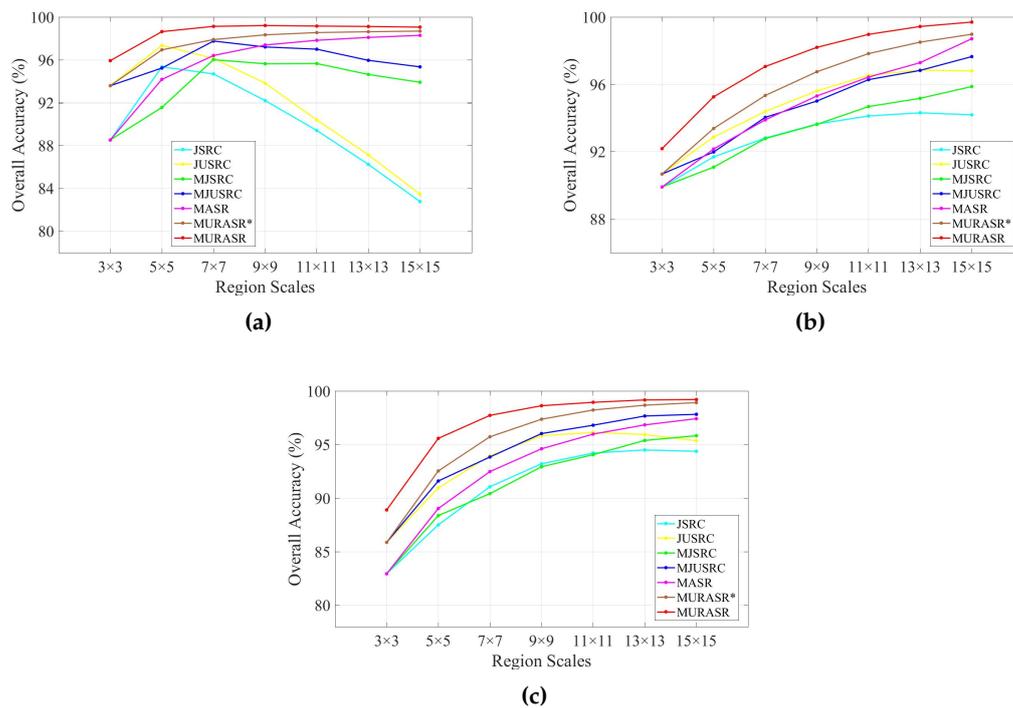
Method	JSRC	JUSRC	MJSRC	MJUSRC	MASR	MURASR*	MURASR
JSRC	–	–16.69	–13.38	–31.89	–28.46	–40.43	–42.27
JUSRC	16.69	–	4.06	–19.51	–11.12	–29.82	–32.77
MJSRC	13.38	–4.06	–	–22.57	–20.97	–32.92	–34.68
MJUSRC	31.89	19.51	22.57	–	5.09	–18.40	–21.43
MASR	28.46	11.12	20.97	–5.09	–	–21.53	–23.41
MURASR*	40.43	29.82	32.92	18.40	21.53	–	–6.93
MURASR	42.27	32.77	34.68	21.43	23.41	6.93	–



**Figure 8.** Classification maps for the University of Pavia image by different algorithms: (a) SRC-Pixel-Wise; (b) JSRC; (c) JUSRC; (d) MJSRC; (e) MJUSRC; (f) MASR; (g) MURASR\*; and (h) MURASR.

#### 4.3. Effects of Region Scales

Except for SRC-Pixel-wise, other related algorithms can be affected by different number of scales. In the previously mentioned experiments, 7 scales have been chosen to compare the performance of all algorithms. The effect of region scales for JSRC, MJSRC, and MASR has been presented in [28].



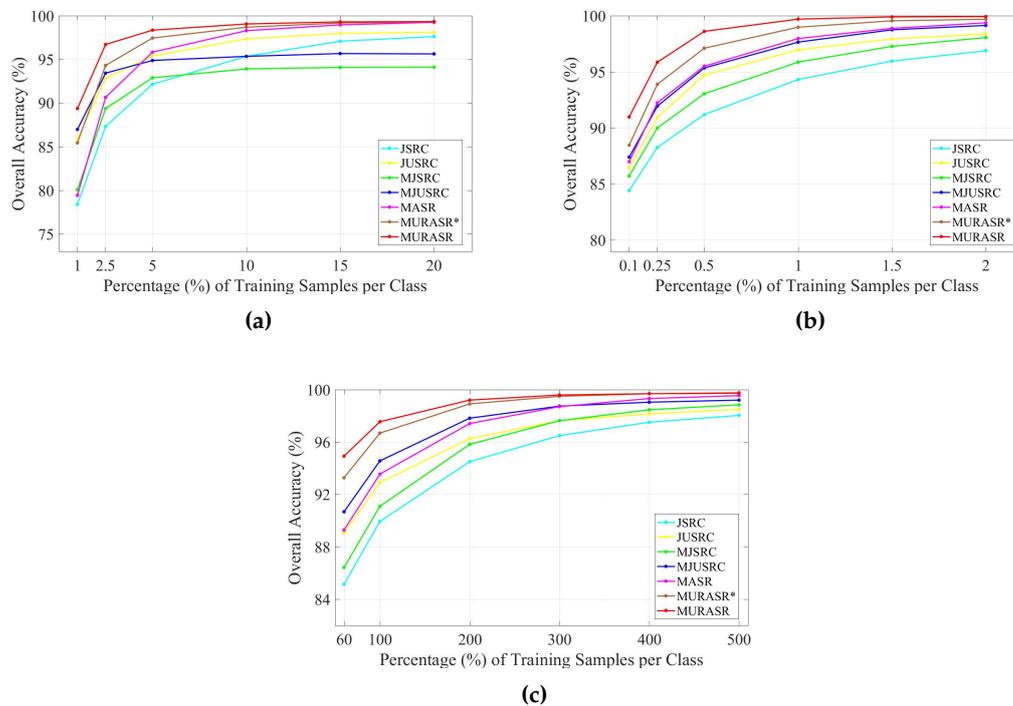
**Figure 9.** Effect of the region scales on single scale algorithms JSRC, JUSRC and the multiscale algorithms MJSRC, MJUSRC, MASR, MURASR\* and MURASR for the: (a) Indian Pine image; (b) Salinas image; and (c) University of Pavia image.

From Table 1, we can find when the scale number is 7, the calculated scale for superpixels is large enough. If the scale continues increasing, there will be more mixed superpixels generated. Moreover, the classification results of MURASR on three images are encouraging when the number of scales is 7. Therefore, the effects of scales number under or equal to 7 will be analyzed in this section. It means that scales for patches range from  $3 \times 3$  to  $15 \times 15$ . Figure 9 shows the average OA of ten runs for JSRC, JUSRC, MJSRC, MJUSRC, MASR, MURASR\* and proposed MURASR. For multiscale algorithms, each scale represents the combination of the current scale and its smaller scales. It is easy to find that the union region based classifiers JUSRC, MJUSRC, and MURASR\* generally outperform corresponding patch region based JSRC, MJSRC and MASR. And the probability majority voting method can optimize the classification result on each region scale. In addition, the proposed MURASR consistently outperforms other algorithms on all the region scales.

#### 4.4. Effects of Training Samples Number

The number of training samples may affect the performance of the classifiers. Therefore the effects of different number of training samples on the JSRC, MJSRC, JUSRC, MJUSRC, MASR, MSPASR and proposed MURASR were examined on the three images. For the Indian Pines, the number of selected training samples for every class varies from 1% to 20% percentage. For the Salinas, the percentage is from 0.1% to 2%. For the University of Pavia, 60–500 training samples were selected for each reference class. The difference in terms of classification OA for each classifier with different number of training samples is illustrated in Figure 10. The OA is also the average of ten runs. As can be observed, the union region based classifiers JUSRC, MJUSRC and MURASR\* always perform better than corresponding patch region based JSRC, MJSRC, and MASR. Comparing the result of MURASR\* and MURASR, it is easy to find that the improvement obtained from probability majority voting

method increases with the decreasing of training samples number. Moreover, the proposed MURASR generally outperforms other classifiers on all the training samples.



**Figure 10.** Effect of the number of training samples on JSRC, JUSRC, MJSRC, MJUSRC, MASR, MURASR\* and MURASR for the: (a) Indian Pine image; (b) Salinas image; and (c) University of Pavia image.

## 5. Conclusions

In this paper, a novel multiscale union region adaptive sparse representation, the MURASR, which uses union region integrating patch and superpixel to exploit the spatial information, is proposed for spectral-spatial HSI classification. Unlike the patch region based MASR, the proposed MURASR extends the patch region to the union region. The union region utilizes the integration of the observation that neighboring pixels that belong to the same material usually are strongly correlated with each other and pixels in the superpixel usually belong to the same material. Before sparse representation, multiscale union regions are generated via the union operation for patch and superpixel. Then multiscale adaptive sparse representation is adopted to classify multiscale union regions and an effective probability majority voting method is applied to generate the final result. Experiments on three HSIs demonstrate that the union region based algorithms always perform better than patch region based algorithms and the proposed MURASR outperforms other algorithms in terms of quantitative metrics and visual quality for the classification maps.

As the MURASR is a pixel-based algorithm, if we replace the superpixel with a region growing up from each test pixel, the generated union region will have more accurate representation of the spatial information. Thus, the further research will generate one superpixel for each test pixel. In addition, the structure dictionary for sparse representation is constructed directly by selected training pixels. A trained structure dictionary may decrease the running time of the algorithm and provide more accurate representation for test pixels.

**Acknowledgments:** This research was supported by the National Natural Science Foundation of China under Grant No. 41171339 and 61501413. The authors would like to thank Ly. Fang for providing the source code of MASR and M.-Y. Liu for over-segmentation methods on their website (<http://www.escience.cn/people/LeyuanFang/index.html>, <http://mingyuliu.net>). The authors would like to thank David A. Landgrebe from Purdue University for providing the AVIRIS image of Indian Pines and Paolo Gamba from University of Pavia for providing the ROSIS data set. The authors would like to thank the National Aeronautics and Space Administration Jet Propulsion Laboratory for providing the AVIRIS image of Salinas. The authors would also like to thank the handling editor and anonymous reviewers for their valuable comments and suggestions, which significantly improved the quality of this paper.

**Author Contributions:** Fei Tong and Hengjian Tong proposed the model and implemented the experiments. Fei Tong wrote the manuscript. Junjun Jiang provided overall guidance of the work and edited the manuscript. Yun Zhang reviewed and edited the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Fauvel, M.; Tarabalka, Y.; Benediktsson, J.A.; Chanussot, J.; Tilton, J.C. Advances in Spectral-Spatial Classification of Hyperspectral Images. *Proc. IEEE* **2013**, *101*, 652–675.
2. Zhang, L.; Zhang, L.; Tao, D.; Huang, X. Sparse Transfer Manifold Embedding for Hyperspectral Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1030–1043.
3. Du, B.; Zhang, L. Random-Selection-Based Anomaly Detector for Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 1578–1589.
4. Zhong, Y.; Wang, X.; Zhao, L.; Feng, R.; Zhang, L.; Xu, Y. Blind spectral unmixing based on sparse component analysis for hyperspectral remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 49–63.
5. Borengasser, M.; Hungate, W.S.; Watkins, R. *Hyperspectral Remote Sensing: Principles and Applications*; CRC Press: Boca Raton, FL, USA, 2007.
6. Plaza, A.; Benediktsson, J.A.; Boardman, J.W.; Brazile, J.; Bruzzone, L.; Camps-Valls, G.; Chanussot, J.; Fauvel, M.; Gamba, P.; Gualtieri, A.; et al. Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.* **2009**, *113* (Suppl. 1), S110–S122.
7. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1778–1790.
8. Bruzzone, L.; Chi, M.; Marconcini, M. A Novel Transductive SVM for Semisupervised Classification of Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 3363–3373.
9. Chi, M.; Bruzzone, L. Semisupervised Classification of Hyperspectral Images by SVMs Optimized in the Primal. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1870–1880.
10. Zhong, Y.; Lin, X.; Zhang, L. A Support Vector Conditional Random Fields Classifier With a Mahalanobis Distance Boundary Constraint for High Spatial Resolution Remote Sensing Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1314–1330.
11. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Semisupervised Hyperspectral Image Classification Using Soft Sparse Multinomial Logistic Regression. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 318–322.
12. Ratle, F.; Camps-Valls, G.; Weston, J. Semisupervised Neural Networks for Efficient Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 2271–2282.
13. Zhong, Y.; Zhang, L. An Adaptive Artificial Immune Network for Supervised Classification of Multi-/Hyperspectral Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 894–909.
14. Ji, R.; Gao, Y.; Hong, R.; Liu, Q.; Tao, D.; Li, X. Spectral–Spatial Constraint Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1811–1824.
15. Khodadadzadeh, M.; Li, J.; Plaza, A.; Ghassemian, H.; Bioucas-Dias, J.M.; Li, X. Spectral–Spatial Classification of Hyperspectral Data Using Local and Global Probabilities for Mixed Pixel Characterization. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6298–6314.
16. Ghamisi, P.; Benediktsson, J.A.; Ulfarsson, M.O. Spectral–Spatial Classification of Hyperspectral Images Based on Hidden Markov Random Fields. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2565–2574.
17. Zhou, Y.; Peng, J.; Chen, C.L.P. Dimension Reduction Using Spatial and Spectral Regularized Local Discriminant Embedding for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1082–1095.

18. Falco, N.; Benediktsson, J.A.; Bruzzone, L. Spectral and spatial classification of hyperspectral images based on ICA and reduced morphological attribute profiles. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6223–6240.
19. Fang, L.; Li, S.; Duan, W.; Ren, J.; Benediktsson, J.A. Classification of Hyperspectral Images by Exploiting Spectral—Spatial Information of Superpixel via Multiple Kernels. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6663–6674.
20. Li, S.; Lu, T.; Fang, L.; Jia, X.; Benediktsson, J.A. Probabilistic fusion of pixel-level and superpixel-level hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7416–7430.
21. Wang, Y.; Song, H.; Zhang, Y. Spectral-spatial classification of hyperspectral images using joint bilateral filter and graph cut based model. *Remote Sens.* **2016**, *8*, 748.
22. Song, H.; Wang, Y. A spectral-spatial classification of hyperspectral images based on the algebraic multigrid method and hierarchical segmentation algorithm. *Remote Sens.* **2016**, *8*, 296.
23. Ma, L.; Ma, A.; Ju, C.; Li, X. Graph-based semi-supervised learning for spectral-spatial hyperspectral image classification. *Pattern Recogn. Lett.* **2016**, *83*, 133–142.
24. Jiang, J.; Chen, C.; Song, X.; Cai, Z. Hyperspectral image classification using set-to-set distance. In Proceedings of the ICASSP, Shanghai, China, 20–25 March 2016; pp. 3346–3350.
25. Jiang, J.; Chen, C.; Yu, Y.; Jiang, X.; Ma, J. Spatial-aware collaborative representation for hyperspectral remote sensing image classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 404–408.
26. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral Image Classification Using Dictionary-Based Sparse Representation. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3973–3985.
27. Yuan, Y.; Lin, J.; Wang, Q. Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization. *IEEE Trans. Cybern.* **2016**, *46*, 2966–2977.
28. Fang, L.; Li, S.; Kang, X.; Benediktsson, J.A. Spectral–Spatial Hyperspectral Image Classification via Multiscale Adaptive Sparse Representation. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 7738–7749.
29. Fang, L.; Li, S.; Kang, X.; Benediktsson, J.A. Spectral–Spatial Classification of Hyperspectral Images with a Superpixel-Based Discriminative Sparse Model. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 4186–4201.
30. Fu, W.; Li, S.; Fang, L.; Kang, X.; Benediktsson, J.A. Hyperspectral Image Classification Via Shape-Adaptive Joint Sparse Representation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 556–567.
31. Zhang, S.; Li, S.; Fu, W.; Fang, L. Multiscale Superpixel-Based Sparse Representation for Hyperspectral Image Classification. *Remote Sens.* **2017**, *9*, 139.
32. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust Face Recognition via Sparse Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227.
33. Leviatan, D.; Temlyakov, V.N. Simultaneous approximation by greedy algorithms. *Adv. Comput. Math.* **2006**, *25*, 73–90.
34. Tropp, J.A.; Gilbert, A.C.; Strauss, M.J. Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit. *Signal Process.* **2006**, *86*, 572–588.
35. Mori, G.; Ren, X.; Efros, A.A.; Malik, J. Recovering human body configurations: combining segmentation and recognition. In Proceedings of the IEEE Conference Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; Volume 2, pp. 326–333.
36. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282.
37. Liu, M.Y.; Tuzel, O.; Ramalingam, S.; Chellappa, R. Entropy rate superpixel segmentation. In Proceedings of the IEEE Conference Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 2097–2104.
38. Felzenszwalb, P.F.; Huttenlocher, D.P. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **2004**, *59*, 167–181.
39. Zhong, Y.; Gao, R.; Zhang, L. Multiscale and multifeature normalized cut segmentation for high spatial resolution remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6061–6075.
40. Vidal, R.; Ma, Y.; Sastry, S.S. Principal Component Analysis. In *Generalized Principal Component Analysis*; Springer: New York, NY, USA, 2016; pp. 25–62.

41. Gualtieri, J.A.; Crompton, R.F. Support vector machines for hyperspectral remote sensing classification. In Proceedings of the SPIE, Washington, DC, USA, 29 January 1999; pp. 221–232.
42. Foody, G.M. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 627–633.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).